BACH2 immunodeficiency illustrates an association between super-enhancers and haploinsufficiency

BACH2 immunodeficiency illustrates an association between super-enhancers and haploinsufficiency

Behdad Afzali[1,2 *†], Juha Grönholm[3 *], Jana Vandrovcova[4,5 *], Charlotte O'Brien[5], Hong-Wei Sun[1], Ine Vanderleyden[6], Fred P Davis[1], Ahmad Khoder[5], Yu Zhang[3], Ahmed N Hegazy[7,8], Alejandro V Villarino[1], Ira W Palmer[1], Joshua Kaufman[1], Norman R Watts[1], Majid Kazemian[9], Olena Kamenyeva[3], Julia Keith[7], Anwar Sayed[5], Dalia Kasperaviciute[10], Michael Mueller[10], Jason D. Hughes[11], Ivan J. Fuss[3], Firas Sadiyah[6], Kim Montgomery-Recht[12], Joshua McElwee[11], Nicholas P Restifo[13], Warren Strober[3], Michelle A Linterman[6], Paul T Wingfield[1], Holm H Uhlig[7,14], Rahul Roychoudhuri[6], Timothy J. Aitman[5,15], Peter Kelleher[5], Michael J Lenardo[3], John J O'Shea[1], Nichola Cooper[5 †‡], Arian DJ Laurence[7,16 ‡]

[1] Lymphocyte Cell Biology Section (Molecular Immunology and Inflammation Branch), Biodata Mining and Discovery Section and Protein Expression Laboratory, National Institutes of Arthritis, and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD, USA. (behdad.afzali@nih.gov; hong-wei.sun@nih.gov; fred.davis@nih.gov; alejandro.villarino@nih.gov; john.oshea@nih.gov; palmeri@helix.nih.gov; kaufmanj@mail.nih.gov; wattsn@mail.nih.gov; wingfiep@mail.nih.gov)

[2] MRC Centre for Transplantation, King's College London, UK. (behdad.afzali@kcl.ac.uk)

[3] Molecular Development of the Immune System Section, NIAID Clinical Genomics Program, Biological Imaging Section (Research Technologies Branch) and Mucosal Immunity Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD USA. (juhapetrikristian.gronholm@nih.gov; mlenardo@niaid.nih.gov; olena.kamenyeva2@nih.gov; yuzhang@mail.nih.gov; ifuss@niaid.nih.gov; wstrober@niaid.nih.gov)

[4] Molecular Neuroscience, Institute of Neurology, Faculty of Brain Sciences, University College London, UK. (jana.vandrovcova@ucl.ac.uk)

[5] Department of Medicine, Imperial College London, UK. (a.khoder09@imperial.ac.uk; anwar.sayed13@imperial.ac.uk; n.cooper@imperial.ac.uk; jana.vandrovcova@ucl.ac.uk; charlotte.obrien06@imperial.ac.uk; p.kelleher@imperial.ac.uk; tim.aitman@ed.ac.uk)

[6] Laboratory of Lymphocyte Signaling and Development, Babraham Institute, Cambridge, UK. (Ine.Vanderleyden@babraham.ac.uk; firas.sadiyah@babraham.ac.uk; rahul.roychoudhuri@babraham.ac.uk; Michelle.Linterman@babraham.ac.uk)

[7] Translational Gastroenterology Unit, Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, UK. (ahmed.hegazy@ndm.ox.ac.uk; arian.laurence@ndm.ox.ac.uk; holm.uhlig@ndm.ox.ac.uk; julia.keith@ndm.ox.ac.uk)

[8] Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK (ahmed.hegazy@ndm.ox.ac.uk)

[9] Departments of Biochemistry and Computer Science, Purdue University, West Lafayette, IN, USA (kazemian@purdue.edu)

[10] Imperial BRC Genomics Facility Hammersmith hospital, Du Cane road, London, UK. (michael.mueller@genomicsengland.co.uk; d.kasperaviciute@imperial.ac.uk)

[11] Merck Research Laboratories, Merck & Co. Inc., Boston, MA, USA (Jason.hughes@merck.com; joshua_mcelwee@merck.com)

[12] Clinical Research Directorate/CMRP, Leidos Biomedical Research Inc., NCI at Frederick, Frederick, MD, USA (kim.montgomery-recht@fnlcr.nih.gov)

[13] National Cancer Institute, National Institutes of Health, Bethesda, MD, USA (restifon@mail.nih.gov)

[14] Department of Paediatrics, University of Oxford, UK (holm.uhlig@ndm.ox.ac.uk)

[15] Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, UK (tim.aitman@ed.ac.uk)

[16] Department of Haematology Northern Centre for Cancer Care, Freeman road, Newcastle upon Tyne, UK. (arian.laurence@nuth.nhs.uk)

[*] These authors contributed equally to this work ‡ These authors contributed equally to this work

[†]Correspondence to: Behdad Afzali (behdad.afzali@nih.gov; behdad.afzali@kcl.ac.uk); Nichola Cooper (n.cooper@imperial.ac.uk).

Abstract

The transcriptional programs guiding lymphocyte differentiation depend on precise expression and timing of transcription factors (TFs). BTB And CNC Homology 2 (BACH2) is a TF essential for both T and B lymphocytes that is associated with an archetypal super-enhancer (SE). Single nucleotide variants (SNVs) in the *BACH2* locus have been associated with multiple autoimmune diseases but *BACH2* mutations have not previously been shown to cause Mendelian monogenic primary immunodeficiency. Herein, we describe an autosomal dominant syndrome of BACH2-related immunodeficiency and autoimmunity (BRIDA) in three patients from two independent kindreds, resulting from haploinsufficiency of *BACH2*. Patients had lymphocyte maturation defects, resulting in both immunoglobulin deficiency and intestinal inflammation. The mutations disrupted BACH2 protein stability by interfering with homodimerization or by causing aggregation. Analogous lymphocyte defects were noted in a murine *Bach2* heterozygous deficiency model, confirming that the loss of one allele is sufficient to recapitulate key cellular phenotypes of the disease. More generally, we found that genes causing monogenic haploinsufficient diseases are substantially enriched for TFs and SE-architecture. These observations show a new feature of genes with SE architecture in Mendelian diseases of immunity and suggest that heterozygous mutations in SE-regulated genes identified on whole exome/genome sequencing may have greater significance than previously recognized.

Introduction

The inheritance pattern of genetic diseases consists of a spectrum, ranging from the vast majority representing polygenic susceptibility variants (usually identified on GWAS studies) to the minority, which are monogenic and manifest in either a recessive or dominant manner. It is now appreciated that mutations in over 300 different genes can cause primary immunodeficiency (PID), many of which affect T and B lymphocyte function[1-4]. PIDs are often paradoxically associated with autoimmunity[3-7]. Common variable immunodeficiency (CVID), a major form of PID with antibody deficiency, is typically associated with recurrent infections and autoimmunity[8]. Recently developed gene-sequencing technologies now allow for rapid identification of PIDs but have also raised the important question of how to interpret the many heterozygous mutations seen in both patients and healthy controls. Relatively few PID syndromes are caused by haploinsufficiency, an autosomal dominant pattern of disease inheritance, where one allele is damaged and only a single functional allele remains[9]. Genes, such as *CTLA4*, are particularly susceptible to haploinsufficiency and the reasons are unknown[10]. In the light of many healthy people that harbor heterozygous loss of function or hypomorphic variants, why should partial changes in gene expression have significant consequences to health?

Promoters and enhancer elements govern gene expression. Most, such as housekeeping genes like actin, are regulated by a limited number of associated enhancers, known as "typical enhancers"[11]. By contrast, 5-10% of genes have a complex enhancer structure consisting of multiple enhancers that collectively are described as SEs[12,13]. Genes with associated SEs have a highly regulated pattern of gene expression; single nucleotide polymorphisms associating in GWAS studies with autoimmune diseases are preferentially enriched within SE regions[14]. These findings suggest the hypothesis that minor changes in regulatory function at SE regions could have significant consequences to the immune system for genes regulated by SEs.

BACH2 is a typical example of an SE regulated gene associated with autoimmune disease. It is a highly conserved member of the basic and leucine zipper domain (bZIP) superfamily of TFs and a critical regulator of both T and B lymphocyte differentiation and maturation[15,16]. Polymorphisms in the human gene locus associate with multiple autoimmune diseases, including asthma[17], insulin dependent diabetes mellitus[18], Crohn's and celiac diseases[19,20], vitiligo[21] and multiple sclerosis[16,22]. The *Bach2* gene locus has the largest SE structure seen in mouse lymphocytes[14]. Homozygous deletion of *Bach2* in mice results in spontaneous fatal autoimmunity between 3 and 9 months of age[15]. Functionally, Bach2 acts as a repressive "guardian" TF that

regulates the balance between a network of other TFs critical to T and B cell specification and maturation. In B cells, Bach2 controls the balance between Pax5 and Blimp1 by repressing the latter[23,24], to decelerate plasma cell differentiation and permit antibody class switch recombination (CSR) (allowing expression of IgA, G and E isotypes)[25]. Consequently, mice lacking Bach2 have B cells with impaired CSR that rapidly differentiate into IgM-restricted plasma cells. In T cells, Bach2 regulates networks of genes that control T cell effector lineages[14] and cellular senescence[26], thus limiting differentiation into effector cells [15] and promoting development of FoxP3$^+$ regulatory T cells (Tregs). Tregs are a non-redundant suppressive lineage of T cells that prevent development of autoimmune diseases by controlling over-activation of the immune system[27]. Thus, mice deficient in Bach2 demonstrate both a paucity of Tregs and an excess of memory/effector T cells that age and die prematurely, resulting in autoimmunity.

Structurally, Bach2 contains a BTB/POZ domain that mediates homo-and hetero-dimerization at its N-terminus and a bZIP domain at the C-terminus required for DNA binding. The dimerization domain is an alpha-helical structure containing a cysteine residue that is capable of forming a disulphide bond with its opposite partner[28]. Thus homo-dimerization is likely to be stabilized by a covalent modification that occurs soon after protein folding. Bach2 dimers translocate to the nucleus where they interact with target DNA loci at palindromic Maf recognition elements (MARE), either alone or in collaboration with other members of the bZIP family, such as the small Maf proteins (MafF, MafG and MafK)[16]. This interaction, for example at the *Prdm1* locus that encodes Blimp1, represses gene expression.

Here we describe a novel PID caused by haploinsufficiency of BACH2 and propose a shared genetic mechanism to explain why some genes are particularly susceptible to causing disease by haploinsufficiency and should be regarded as significant and prioritized in the interpretation of heterozygote variants found on whole exome sequencing.

Results

Mutations in the *BACH2* gene locus are associated with CVID and colitis

We investigated a female (**Figs. 1a and 1b** – Family A) with infancy-onset colitis, who became ill at 19 years old with non-infectious fever, splenomegaly (21.7 cm, compared to 10-12 cm in normal adults) (**Fig. 1c**) and pancytopenia. Fever and cytopenia improved with corticosteroids, but lymphopenia, deficiency in immunoglobulin (Ig)M, IgG, IgA and IgE, ongoing colitis, lung infiltrates and recurrent upper respiratory tract infections persisted (**Fig. 1c and Table 1 and Supplementary Table 1**). A colonic biopsy demonstrated inflammatory changes with crypt branching and prominent lymphocytic infiltrates around the crypts (**Fig. 1d**), with significantly reduced $FoxP3^+$ regulatory T (Treg) cells compared with healthy controls or patients with classical IBD (**Fig. 1e**). The early disease onset and unusual symptoms in the absence of family history prompted us to perform whole exome sequencing on the patient and healthy parents as a trio. After excluding all variants with minor allele frequency (MAF) >0.01, no candidate variants remained to support a hypothesis of recessive inheritance. We found a heterozygous novel *de novo,* non-synonymous mutation in *BACH2,* c.T71C, predicted to be deleterious (**Supplementary Table 2**), substituting a highly conserved leucine with proline (L24P), and not present in healthy family members (**Fig. 1b and Supplementary Figure 1**). A second family that had been previously investigated by exome sequencing (**Fig. 1a**– Family B) was found to have a heterozygous point mutation in *BACH2*, c.G2362A (**Fig. 1b**), substituting glutamic acid with lysine (E788K) in a father and daughter, both of whom presented with inflammation of both small and large bowel, together with pulmonary disease, including recurrent sino-pulmonary infections, bronchiectasis and fibrosis (**Fig. 1c and Supplementary Table 1**). The *BACH2* mutation was not seen in healthy family members (**Supplementary Figure 1**). The father (proband) was deficient in all Ig sub-types; his daughter had undetectable IgA (**Supplementary Table 1**). Detailed clinical features are described in the **Supplementary text, Tables 1 and Supplementary Fig 1**. We found no low MAF variants nor causative mutations in genes causing monogenic IBD or other recognized primary immunodeficiencies[29-31].

In the lymphocytes of affected individuals, we found decreased expression of $FoxP3$[27] in $CD4^+CD25^{high}CD127^{low}$ regulatory T cells (Tregs) (**Fig. 2a**) and increased expression of the Th1 transcription factor T-bet and two gut-homing receptors, CCR9 and β7-integrin on $CD4^+$ T cells[32,33] (**Fig. 2b**). In the patient B cells, we found a marked reduction in $CD19^+CD27^+$ memory and IgG class-switched $CD27^+IgG^+$ B cells (**Fig. 2c**). These features were not present in healthy controls or patients with inflammatory bowel disease (IBD) (**Supplementary Fig. 2a**).

Furthermore, $CD24^+CD38^+$ transitional B cells were increased in patients (**Supplementary Fig. 2b).** *In vitro* activation of naïve B cells from patients resulted in significantly impaired plasmablast generation, class-switch recombination and class-switched antibody secretion in the presence of IL-21 (**Figs. 2d and 2e**), suggesting a defect in B cell maturation towards memory and plasma cells, similar to *Bach2* knockout mice[34]. Polyclonal activation of T cells resulted in reduced $CD4^+$ T cell proliferation compared with healthy controls (**Supplementary Fig. 2c**). In summary, the immunopheontype of patients with mutations in *BACH2* consisted of compromised Tregs, enhanced Th1 differentiation, impaired proliferation and defective B cell maturation and Ig class switching.

BACH2 gene mutations result in a reduction in BACH2 protein in patients and cell lines

We next measured BACH2 protein expression by flow cytometry and found it was reduced in patient $CD4^+$, $CD8^+$ and B lymphocytes despite normal mRNA expression (**Figs. 3a and 3b**). We measured protein expression of Flag-tagged vectors encoding wild-type (WT) or mutant forms of *Bach2* in transfected HEK293T cells and found that mutant forms of the protein accumulated less than WT (**Fig. 3c**), at all time points measured and concentrations of vector used (**Supplementary Fig. 3a-b**). *PRDM1*, which encodes the protein BLIMP1, is a target of BACH2-mediated transcriptional repression[24]. We found that patient naïve B cells and $CD4^+$ T cells expressed significantly higher levels of *PRDM1* mRNA compared with healthy controls suggesting a release from BACH2 repression (**Figs. 3d and 3e**)**.** Furthermore, this difference could be reversed by forced expression of WT BACH2 in patient $CD4^+$ lymphocytes **(Fig. 3e**). These observations suggested a causal relationship between reduced BACH2 levels in patients and their cellular phenotype. To confirm this relationship, we silenced BACH2 expression in healthy control T and B cells using RNAi by ~50% and carried out functional phenotyping (**Supplementary Figs. 4a and 4b**). Silencing BACH2 in control $CD4^+$ T cells led to a significant rise in *PRDM1* mRNA (**Fig. 3f**) and resulted in a reduction in proliferation of $CD4^+$ T cells (**Supplementary Fig. 4c**), in similar fashion to that seen in primary $CD4^+$ T cells of patients (**Supplementary Fig. 2c**)**.** In addition, silencing BACH2 in healthy control B cells, significantly suppressed *in vitro* class switch recombination towards the IgG and IgA isotypes (**Fig. 3g**). Thus, experimental silencing of BACH2 in healthy T and B cells recapitulated the phenotype seen in primary cells of the patients.

Reduced BACH2 protein expression is caused by the production of abnormal proteins

Both mutations that we identified affect highly conserved amino acid residues in BACH2 (**Fig. 4a**). Murine and human BACH2 share 90% sequence identity and L24 is conserved across species and with other members of the BTB/POZ domain family (**Supplementary Figs. 5a and 5b; Supplementary Table 2**). L24 resides within α-helix-1 (residues 18-34) of the BTB/POZ domain, a key part of the BACH2 homo-dimerization interface (**Figs. 4b and 4c**). The mutant proline residue likely perturbs α-helix-1 of the BTB/POZ domain and places a polar residue into the hydrophobic face of that helix, which we predicted would decrease dimer stability (**Supplementary Table 2**). We expressed and purified the BTB domains from both WT and L24P mutant proteins. The WT protein was soluble and formed dimers (**Fig. 4d**), whereas the L24P mutant was insoluble in solution, likely misfolded, and formed multiple aggregated species (**Fig. 4e**). E788, the site of the C-terminus mutation, is again highly conserved (**Supplementary Fig. 5a)**. Though not characterized by structural studies, it is in proximity to a nuclear export signal (**Fig. 4a**). We found that wild type BACH2 protein was evenly distributed in both cytoplasm and nucleus, whereas the E788K mutant protein was aggregated in the cytoplasm with relatively little in the nucleus (**Fig. 4f and Supplemental Movies 1-2**). Similar protein aggregates were observed in HEK293T cells transfected with this C-terminal variant (**Supplementary Fig. 5c and Supplemental Movies 3-4**). By contrast, aggregates were not detected in lymphocytes expressing the L24P mutant from patient A.II.1, although, as noted, L24P mutant BACH2 protein expression levels were lower than WT control (**Supplementary Fig. 5d**).

***BACH2* gene mutations do not confer a dominant negative effect**

In both families, the *BACH2* gene mutations could potentially act in a dominant negative manner. To test this, HEK293T cells were co-transfected with Flag-tagged WT together with untagged WT or mutant Bach2 protein-expressing constructs. Neither patient mutant altered WT protein expression (**Fig. 5a**). The experiment was repeated with HEK293T cells co-transfected with vectors encoding two tagged WT (HA-Bach2 and Flag-Bach2) forms of the proteins together with either untagged WT or mutant Bach2 protein-coding constructs (**Fig. 5b**). Co-immunoprecipitation studies showed that WT untagged Bach2, but not mutant forms of the protein, interfered with dimerization between HA and Flag-tagged WT Bach2. Furthermore, when WT Flag-Bach2 was co-transfected together with HA-tagged WT, L24P or E786K Bach2, we detected reduced mutant HA-Bach2 bound to Flag-tagged WT Bach2 after immunoprecipitation, in proportion to the reduction in protein accumulation, implying limited, if any, effects on WT Bach2 (**Supplementary Fig. 6a)**. All these results were consistent with our

earlier findings of loss of stability of the mutant proteins compared with wild type proteins (**Fig. 4**). Finally, we used retroviral constructs encoding murine WT or mutant *Bach2* genes to transduce *Prdm1*-YFP transgenic $CD4^+$ T cells. Forced expression of wild type Bach2 alone led to a significant reduction in the expression of *Prdm*1-YFP, but co-transduction with either mutant form of *Bach2* did not interfere with repression of the *Prdm1* reporter in primary mouse lymphocytes (**Fig. 5c**). Collectively, these data indicate that neither *BACH2* mutation exerted a dominant negative effect.

***Bach2*$^{+/-}$ mice have similar defects in T and B cells**

In the absence of a dominant negative effect we next turned to haploinsufficiency as an explanation. Complete absence of *BACH2* in mice results in B cell immunodeficiency and fatal autoimmunity later in life[15,16]. If haploinsufficiency is responsible for the defects in lymphocyte development observed in our patients, we would expect to see a similar effect in mice heterozygous for WT and null alleles (*Bach2*$^{+/-}$). We found that *Bach2*$^{+/-}$ animals manifest reduced *Bach2* mRNA (**Fig. 6a**) and protein expression **(Fig. 6b**) together with elevated *Prdm1* mRNA **(Supplementary Fig. 7a**). There was no difference in the numbers of $CD4^+$ and $CD8^+$ T cells, B cells or plasma cells in unchallenged mice **(Supplementary Figs. 7b and 7c**) but *Bach2*$^{+/-}$ mice did have a small but significant reduction in $Foxp3^+$ cells together with significant increases in $CCR9^+$ and β7-integrin$^+$ cells in $CD4^+$ T cells **(Figs. 6c and 6e)**. We next immunized WT and *Bach2*$^{+/-}$ mice with 4-hydroxy-3-nitrophenylacetyl hapten-conjugated chicken gamma globulin (NP-CGG) in alum and analyzed the splenic B cell response. Immunized *Bach2*$^{+/-}$ mice exhibited minimal induction of both IgG1 class switched-$B220^{high}CD138^-$ B cells and $B220^{low}CD138^+$ plasma cells compared to WT mice (**Fig. 6f**). The proportion of germinal center $B220^+Ki67^+Bcl6^+$ B cells was also reduced in *Bach2*$^{+/-}$ mice (**Fig. 6g**), supporting a haploinsufficiency model.

Genes that are associated with super-enhancer structures are associated with haplo-insufficiency syndromes.

Taken together, our data argue that the maintenance of a threshold concentration of BACH2 is crucial for proper immunoregulation. Mutations of other TFs have been reported to cause haploinsufficient disorders[35]. *BACH2* expression is regulated in a complex manner and the *BACH2* locus contains an archetypal SE (**Fig. 7a)**[12-14,16,36,37]. We therefore hypothesized that SE structure may be enriched among genes causing haploinsufficiency (HI) diseases. To this end, we compared genetic disorders mediated by HI (372 genes) versus autosomal recessive (AR)

inheritance (259 genes) to haplo*sufficient* (HS) genes (those where single allele deletions are inconsequential; 901 genes) (Ref. [38] and Methods). To validate these three groups, we evaluated the probability of loss of function intolerance (pLI) score (as estimated by ExAc[39]), where a score of 0 predicts that loss of a single copy of the gene is well tolerated whereas a score of 1 predicts that loss of a single copy is poorly tolerated and likely to result in a disease. As expected, the median pLI score for our HI list was significantly higher than the others (median values of 0.86, 0.0005 and 0.004 for HI, HS and AR recessive genes, respectively) (**Fig. 7b**). Moreover, HI genes were substantially more likely to have SE architecture, as denoted by especially high acetylated histone H3 lysine 27 (H3K27Ac) signal, a hallmark of active enhancers[40] (**Figs. 7c and 7d, Supplementary Fig. 8a** and **supplementary Table 3**). In contrast, there was no difference in the frequency of typical enhancers between the three groups (**Fig. 7d**). We next compared the function of genes between the three groups and found that HI genes were more likely to encode transcription factors than genes associated with AR inheritance or HS genes (**Fig. 7e, Supplementary Figs. 8b and 8c**). To address any potential confounding abundance of transcription factor genes in SE, we also divided our list of HI genes into those that code for transcription factors and those that code for all other proteins and compared the frequency of SEs (**Supplementary Fig. 8c**). We found that even after discounting TF genes, haploinsufficiency disease-causing genes are heavily enriched for SE architecture compared to HS and AR genes (**Supplementary Fig. 8d**). We next asked whether SE-bearing genes have lesser tolerance to loss-of-function mutations and whether the 'size' of the SE correlates with this effect. We expanded our analysis to a collection of genes regulated by SEs called across more than 100 tissues (dbSuper database[41]) and observed both a striking increase in the probability of loss of function intolerance score with increasing SE signal size and a concomitant increase in the proportion of transcription factor genes (**Fig. 7f**). Thus, not only the presence of an SE but also its "size" correlates with likelihood of disease caused by haploinsufficiency.

SE architecture has been shown previously to associate with human disease loci in genome-wide association studies (GWAS)[14,36,42]. This is the case for *BACH2*[16,22], which was consistently in the top 1% of human SE genes by H3K27Ac SE signal intensity in naïve $CD4^+$ T, naïve $CD8^+$ T and B cells (**Supplementary Figs. 8e-g**). Based on the SE enrichment among HI genes, we next asked whether there would be general enrichment of GWAS "hits" in genes associated with haploinsufficient disease. In agreement with this hypothesis, we found that there was highly significant enrichment of disease-associated SNVs within this gene set (**Fig. 7g** and **Supplementary Table 5**). To exclude gene size as a potential confounding factor, the analysis

was repeated on subsets of genes of less than 50 kb and again we found more GWAS associations in genes associated with HI syndromes compared to HS genes (**Supplementary Fig. 8h**). Thus, HI genes are enriched for both SEs and GWAS "hits".

Discussion

Adaptive immunity is critically dependent on appropriate differentiation and maturation of lymphocytes. Several complex steps of differentiation are required to form mature cells that occupy specific niches and carry out defined roles within the immune system. Key to the regulation of lymphocyte differentiation is precise control over expression of many transcription factors (TFs) that form complex regulatory networks. The identification of both mice and humans with dramatic early onset stereotypical autoimmune disease associated with a homozygous loss of gene expression has led to the identification of many key regulatory TFs, most notably FOXP3, the master TF of T regulatory (Treg) cells[27].

The reduction in the cost and time it takes to perform whole exome sequencing has allowed patients with no family history to be analysed for genetic mutations. Comparing patients' DNA sequence with healthy parents identifies the appearance of *de novo* mutations that would otherwise be missed if a positive family history was required prior to any investigation. Using this strategy a number of heterozygous mutations associated with autoimmune diseases have recently been discovered.

BACH2 plays a major role in the regulation of the adaptive immune system. Its own expression is tightly regulated by the presence of a large super-enhancer region within the *Bach2* locus[14]. The role of BACH2 has been elucidated by the investigation of *Bach2* deficient mice that have a defect in B cell class switch recombination together with a deficiency of Treg differentiation. In mice, this combination results in a chronic variable immunodeficiency together with a late onset, but progressively fatal, autoimmune syndrome that includes inflammatory enteropathy and respiratory infiltrates[15]. In keeping with its place as an SE associated gene, there is a link between single nucleotide polymorphisms within the *BACH2* locus and a number of autoimmune/inflammatory diseases.

Herein we describe three patients from two families that have heterozygous mutations in *BACH2*. Two of the three presented with a history of early onset autoimmune gastrointestinal disease and the third presented later in life. All three have developed a chronic variable immunodeficiency characterized by recurrent respiratory tract infections associated with an inability to generate appropriate antibody responses to vaccination. Our findings support a role for human BACH2 as a key regulator of the human adaptive immune system critical to maintain Treg cell function and B cell maturation. *Bach2* deficient mice exhibit accelerated T cell senescence[26,43] and, in keeping

with this, T cells from our patients have a defect in cell proliferation associated with a progressive T cell lymphopenia. Many of the autoimmune phenomena in our patient with the L24P mutation have been successfully treated with corticosteroids although this has not reduced her chronic variable immunodeficiency nor her pneumonitis, which is of some concern as this is a key cause of early mortality in Bach2 deficient mice. The father with the E788K mutation has developed bronchiectasis later in life. It remains to be seen whether the pneumonitis will be progressive in our L24P patient and result in chronic lung damage.

In the first family, the mutant $BACH2^{T71C}$ gene resulted in a protein that is predicted to be unable to dimerize and is unstable. In the second family, the mutant $BACH2^{E788K}$ protein again showed some evidence of a defect in stability but this was less dramatic, and we saw more evidence of a defect in the localisation of the protein with reduced nuclear localisation. We found little evidence that either mutant protein acted in a dominant negative manner. Thus we attribute the clinical phenotype to *BACH2* HI and this conclusion was consistent with our findings and previous reports[44] that $Bach2^{+/-}$ heterozygote have defects in CSR antibody responses.

Mammalian cells contain tens of thousands of gene enhancer sites that cluster in large numbers around a select subset of genes that make up some 5-10% of the total human genome. These clusters are collectively known as SEs. GWAS associated mutations tend to associate with these gene loci but the significance of this remains unclear. Previous work would suggest that SE genes code for proteins whose function is highly dependent on transcription, small changes of which would lead to significant changes in cell development. From this we hypothesise that SE genes would be susceptible to gene dosage effects in patients with heterozygous mutations. We conclude that the relationship between GWAS studies and SE regulated genes occurs not simply because these genes transcribe proteins that are important *per se* but because small changes in the expression of SE genes result in large functional changes in the affected cells.

In summary, we describe a new disorder, BACH2-related immunodeficiency and autoimmunity (BRIDA) due to heterozygous mutations in *BACH2*. We found that the mechanism of disease is *BACH2* HI and that *BACH2* is a prototype HI gene exhibiting SE architecture. Given the prevalence of heterozygous variants in non-consanguineous human genomes[45], it is difficult to predict which ones cause disease. We demonstrate that HI diseases are associated with heterozygous mutations in SE-regulated genes. As SEs allow complex regulation of gene transcription, we conclude that HI genes are carefully regulated due to their SE association and

that small changes in their expression level can potentially lead to amplified changes in their associated network, especially for TF genes, resulting in significant pathology. Thus, SE-regulated genes should be more significantly prioritized when interpreting heterozygous variants discovered on whole exome/genome sequencing.

Methods

Detailed materials and methods are given in the supplementary text.

Acknowledgments: We thank the patients and healthy donors for their support and Helen Matthews and Clare Neurwirth for coordinating control blood samples. This research was supported by the Intramural Research Programs of NIAMS, the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, Clinical Center, and National Human Genome Research Institute, National Institutes of Health. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This work was supported by Crohn's & Colitis Foundation of America (A.L., H.H.U.), Sigrid Juselius and Emil Aaltonen Foundations (both J.G.), Wellcome Trust (B.A. and R.R.), European Molecular Biology Organization (A.N.H.), a Marie Curie fellowship (A.N.H.), Imperial College National Institute for Health Research (NIHR) Biomedical Research Centre (N.C. and P.K.), Oxford NIHR Biomedical Research Centre (H.H.U.), Chelsea & Westminster Hospital Charity (C.O'B.), UK Biotechnology and Biological Sciences Research Council (R.R), Westminster Medical School Research Trust (P.K), Biotechnology and Biological Sciences Research Council (M.A.L and I.V) and Cambridge Trust (I.V), Leona M. and Harry B. Helmsley Charitable Trust and ESPGHAN (H.H.U.), the MRC Clinical Sciences Centre (CSC) (T.J.A.) and by the CSC Genomics Core Laboratory and by MRC transition funding (T.J.A.). We acknowledge the contribution of the BRC Gastrointestinal biobank/Oxford IBD cohort study, which is supported by the NIHR Oxford Biomedical Research Centre. We thank G. Vahedi, E. Mathé, S. Parker, C. Kanellopoulou and S. Muljo for critically reading the manuscript, J. Kabat for his help on confocal image analysis and S.S. De Ravin and H. Malech for their advice in the use of MaxCyte. Molecular graphics and analyses were performed with the UCSF Chimera package, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). This study utilized high-performance computational capabilities of Helix Systems at the NIH, Bethesda, MD (http://helix.nih.gov).

Author contributions: B.A., J.G. and J.V. designed and performed experiments, analyzed data and wrote the manuscript. C.O'B., I.V., F.P.D., A.K., A.N.H., J.Ke., A.S., R.R., M.L., O.K., H-W.S., Y.Z. performed experiments and/or analyzed data. I.J.F., W.S., T.J.A., P.K., N.C. provided patient samples and clinical and scientific input. K.M-R. co-ordinated patient samples. Patient sequencing and sequence analysis was carried out by J.V., N.C., T.J.A., D.K., M.M., J.D.H., J.McE. and Y.Z. A.V.V., N.W., H.H.U., M.K. provided scientific input. P.T.W. I.W.P., J.Ka. provided scientific input, performed protein chemistry experiments and analyzed data. N.P.R. provided murine reagents for these experiments. M.J.L., J.J.O'S., N.C and A.D.J.L provided scientific input, supervised the project and wrote the manuscript.

Competing financial interests: The authors have no competing interests to declare. Unrelated to this project, H.H.U. declares industrial project collaboration with Lilly, UCB Pharma and Vertex Pharmaceuticals. Travel support was received from Actelion, and MSD.

Figure legends

Figure 1. **Pedigrees and phenotype of patients with mutations in *BACH2***. (**a**) Pedigrees of two families with heterozygous missense coding mutations in *BACH2*, resulting in L24P (left) and E788K (right) amino acid substitutions. Shown are affected heterozygotes (filled symbols) and unaffected family members (open symbols). Arrows indicate probands; WT = wild-type allele; Mut = mutant allele. (**b**) Sanger sequencing chromatograms of the affected individuals in both families. For each individual, the two alleles of the sequenced region of *BACH2* and base positions are shown above the chromatograms. Subject A.II.1 had a heterozygous T to C mutation at coding position 71 whereas patients B.II.1 and B.III.2 were heterozygous for G to A base substitutions at position 2362. (**c**) Computerized tomography scans showing splenomegaly (arrow in upper left) and pulmonary nodules (red circle in upper right) in patient A.II.1 and bronchiectasis (dilated airways; arrow in lower left) and fibrosis ("honeycombing" circled in lower right) in subject B.II.1. (**d**) Photomicrograph of a hematoxylin and eosin-stained section from a colonic biopsy from patient A.II.1 showing crypt branching and lymphocytic inflammatory infiltrate around the crypts. (**e**) Immunofluorescent staining of colonic biopsy from patient A.II.1, control IBD patient and healthy control for nuclear DNA (DAPI, blue), CD3 (green) and FoxP3(orange). Shown are representative sections (left) and cumulative (mean ± sem) quantification (right) from four low power fields per patient (500-3000 $CD3^+$ cells counted per low power field); white scale bar = 100 μm in main image and 2 μm in insets. *$p<0.05$, **$p<0.01$ by t-test.

Figure 2. Immunophenotype of patients with mutations in *BACH2*. (**a-c**) Treg (**a**), T cell (**b**) and B cell (**c**) immunophenotype of patient and healthy control peripheral blood cells. Shown are total FoxP3 expression (mean fluorescent intensity (MFI)) within $CD4^+CD25^{hi}CD127^{lo}$ cells (**a**), expression of the transcription factor T-bet and gut-homing receptors (CCR9 and β7-integrin) in bulk $CD4^+$ T cells (**b**), total memory (**c, left**) and class-switched memory (**c, right**) in bulk B cells. (**d-e**) Plasmablast formation (**d**, left panels), IgG class switch recombination (**d**, right panels) and immunoglobulin secretion (**e**) in naïve patient and healthy control B cells activated *in vitro* as indicated. Shown are representative flow cytometry plots and cumulative data. N.D. = not detected; very low values are shown above the bars for clarity. In (**a-d**) representative flow cytometry plots are shown together with cumulative data (mean ± sem; n = 6 from two independent experiments (3 patients and 3 controls measured twice)) from all patients and matched controls. Note that IgG secretion in (**e**) does not include patient B.III.2, who has normal

IgG secretion. Bars show mean ± sem throughout. *p<0.05 **p<0.01 ***p<0.001 by t-test (**a-c**) and ANOVA (**d** and **e**).

Figure 3. The cellular phenotype is attributable to reduced *BACH2* protein expression. (**a**) BACH2 protein expression in primary immune cells of patients and controls. Shown are representative flow cytometry plots with MFIs indicated (left panels) and cumulative BACH2 protein expression (right panels) from patients relative to controls. (**b**) Cumulative *BACH2* mRNA expression from naïve B cells of patients and controls. (**c**) Representative western blot for Flag and Hsp70 from lysates of HEK293T cells transfected with empty vector (EV), Flag-tagged WT or mutant murine *Bach2* (L24P or E786K, the murine equivalent of E788K). Shown are a representative blot (left) and cumulative quantifications from n = 5 experiments (right). (**d**) *PRDM1* mRNA expression in naïve B cells from patients and healthy controls: cumulative data. (**e and f**) *PRDM1* mRNA expression in $CD4^+$ T lymphocytes of healthy controls and patients transfected with either control or *BACH2* (**e**) and healthy donor $CD4^+$ T lymphocytes transfected with control or BACH2 RNAi (**f**). (**g**) Plasmablast formation, IgG class switch recombination and IgA secretion in naïve healthy control B cells transfected with control RNAi or RNAi specific for BACH2 and activated *in vitro* as shown (n = 5 independent experiments). Shown are representative flow cytometry examples and cumulative data. In (**a-g**) bars show mean values ± sem; patient data are n = 6 from two independent experiments (3 patients and 3 controls measured twice); *p<0.05, **p<0.01, ****p<0.0001 by t-test (**a, d and f**) and ANOVA (**c, e** and **g**).

Figure 4. *$BACH2^{L24P}$* haploinsufficiency is caused by failed protein dimerization. (**a**) Schematic of BACH2 protein with indicated domains and sites of point substitutions in patients. BTB/POZ, BR-C, ttk and bab or Pox virus and Zinc finger domain; bZIP, basic leucine zipper; NES, nuclear export signal. (**b**), The structure of the Bach2 POZ domain. Ribbon representations are based on the crystal structure of the POZ domain, form II (PDB: 3OHV). The structure of the wild type protein is shown together with an expanded and rotated view of the interface. The intermolecular disulfide formed at position 20 is shown in yellow. The leucine residues at position 24 are shown in orange. (**c**), (top) Interface of the WT POZ domain dimer (PDB: 3OHV); (bottom) an homology model of the $BACH2^{L24P}$: WT POZ domain hetero-dimer, illustrating local changes. In each model, one monomer has been rendered as a partially transparent hydrophobicity surface (Orange = hydrophobic, White = intermediate, Blue = hydrophilic). The other monomer is represented as a ribbon (Green). The side chains of selected residues at the dimer interface are shown as sticks. Cys20 (Yellow) and Ile23, Leu24, and Leu27 (all shown in Orange) form a hydrophobic patch on α-helix-1, and two of these patches are in

close contact with one another at the WT dimer interface. Please note that the lower diagram is not meant to represent the structure accurately but is shown merely to indicate regional changes. (**d-e**) Analytical ultracentrifugation carried out on purified wild type (WT) p.BACH2 (**d**) and mutant p.BACH2^{L24P} (**e**) BTB/POZ domain. The direction of sedimentation is left to right and M indicates the sample meniscus. The WT protein is dimeric (35 kDa) as determined by sedimentation equilibrium measurements (shown in **d,** right) and migrates with a single boundary with a sedimentation coefficient (S) of 2.6. The mutant exhibits several boundaries with S values ranging from 4 to 18, indicating heterogeneous large protein aggregates (**e**). (**f**) Confocal microscopy images of primary lymphocytes from healthy control and patient B.II.1 stained for BACH2 (green) and Hoechst (blue). Insets show enlarged images of BACH2 staining; cytoplasmic aggregates in the patient are highlighted (white arrows). Scale bars: 5 μm in main image and 2 μm in inset. Images are representative from 3 independent experiments. Bar graphs quantify the number of cells containing aggregates per high power field (HPF) and nuclear localization of BACH2. *p<0.05 by t-test.

Figure 5. Mutant forms of *Bach2* do not exert dominant negative effects. (**a**) Western blot for Flag and Hsp70 in HEK293T cells co-transfected at 1:1 ratio with Flag-tagged WT murine Bach2 and untagged WT and mutant forms of murine Bach2. Shown is a representative from n = 3 independent experiments. (**b**) co-immunoprecipitation of Flag- and HA-tagged WT *Bach2* transfected into HEK293T cells together with untagged WT and mutant forms of murine Bach2 at 1:1:1 vector ratio. Shown is a representative example from n = 3 independent experiments (left) and quantification of the co-immunoprecipitated Flag and HA signals (right). (**c**) Blimp1-YFP signal in Blimp1-YFP Tg mouse CD4$^+$ T cells co-transduced at 1:1 ratio with retrovirus supernatants encoding WT and mutant forms of murine Bach2. Shown is a representative example (left) and cumulative data (mean ± sem) from n = 4 independent experiments (right). ****p<0.0001 by ANOVA.

Figure 6. *Bach2* haploinsufficient mice have abnormal B cell differentiation and Treg numbers. (**a**) Expression of *Bach2* mRNA in B cells of *Bach2*$^{+/+}$ and *Bach2*$^{+/-}$ mice (n=3 per genotype). (**b**) Bach2 protein expression in splenic naïve B cells of *Bach2*$^{+/+}$ and *Bach2*$^{+/-}$ mice. Shown is a representative example (left) and cumulative quantification (mean ± sem) (right) from n=3 independent experiments. (**c-e**) Flow cytometry analysis of CD4$^+$ splenocytes in *Bach2*$^{+/+}$ and *Bach2*$^{+/-}$ mice. Percentage Foxp3$^+$ (**c**), CCR9$^+$ (**d**) and β7-integrin$^+$ (**e**) are shown with representative flow cytometry plots together with histograms (mean ± sem) generated by measuring a minimum of six mice per group. (**f**) IgM and IgG1 staining of B cells (upper panels)

and plasma cells (lower panels) in splenocytes of *Bach2*$^{+/+}$ and *Bach2*$^{+/-}$ mice 8 days following immunization with 4-Hydroxy-3-nitrophenylacetyl hapten-conjugated chicken gamma globulin (NP-CGG) in alum. Representative flow cytometry plots together with histograms (mean ± sem) generated by measuring seven mice per group. (**g**) B220^{+}Ki67^{+}Bcl6^{+} germinal center B cells in splenocytes of *Bach2*$^{+/+}$ and *Bach2*$^{+/-}$ mice 8 days after immunization with NP-CGG in alum. Representative flow cytometry plots together with histograms (mean ± sem) generated by measuring seven mice per group. All data are representative of 2 independent experiments. *p<0.05, **p<0.01, ***p<0.001 by one-way ANOVA (**f**) and Mann-Whitney U-test (all other panels).

Figure 7. Coding mutations in genes regulated by super-enhancer (SE) structures cause haploinsufficient diseases in humans. (**a**) The *BACH2* locus has SE structures in multiple human immune cell types demarcated by H3K27Ac loading. Red fill denotes the presence of an SE in the *BACH2* locus in a tissue. Source data are indicated. (**b**) Violin plots showing probability of loss of function intolerance scores in haplosufficient (HS), autosomal recessive (AR) and haploinsufficient (HI) gene sets. The white circles show median values. Source data: ExAc database[39]. (**c**) Number of HS, AR or HI genes with and without associated SE architecture in humans (see also **supplementary Fig. 8a and supplementary Table 3**). (**d**) Pie charts indicating the frequency of SE (upper panels) and typical enhancer (TE; lower panels) structures in HS (left), HI (middle) and AR (right) genes. (**e**) Gene ontology (GO) functional annotation enrichment in HI genes. Shown are enrichment scores (blue bars) and Benjamini p-values (in orange) for the top 5 most significantly enriched terms. (**f**) Median probability of loss of function intolerance (black line) against SE signal size; the percentage of genes that are transcription factors (TF, red line) against SE signal size is shown in the inset. For reference, the red line asymptotes to the expected level (mean percentage of genes in the human genome that are TFs is 7.5%). Source data: ExAc[39] and dbSuper[41] databases. (**g**) Pie charts indicating the percentage of HS or HI genes that have GWAS disease associations. p-values in **d** and **g** are Fisher exact tests; NS = non-significant; GWAS = genome-wide association study.

Table 1. Summary clinical characteristics of patients with missense mutations in *BACH2*. IvIg, intravenous immunoglobulin; EBV, Epstein-Barr virus; RhF, rheumatoid factor, dsDNA, double-stranded DNA; ANCA, anti-neutrophil cytoplasmic antibody; p-ANCA, perinuclear ANCA; ANA, antinuclear antibody; UC, ulcerative colitis; N/A not assessed. † Absolute values given in Supplementary Table 1; * progressive decline in IgG; ‡ positive by immunofluorescence but negative for myeloperoxidase and proteinase III antibodies by ELISA.

	Patients		
Demographic and clinical characteristics	**A.II.1**	**B.II.1**	**B.III.2**
Age, Sex	19, F	63, M	40,F
Lymphadenopathy	Yes	Yes	Yes
Splenomegaly	Yes	No	No
Intestinal manifestations	Yes	Yes	Yes
Chronic diarrhea	Yes	Yes	Yes
IBD	Colitis	Not biopsied	UC aged 10; Crohn's aged 32
Pulmonary manifestations	Yes	Yes	Yes
Recurrent sino-pulmonary infections	Yes	Yes	Yes
Radiographic changes on chest CT	Yes	Yes	Not imaged
Immunoglobulins†			
IgM	Low	Low	High
IgG	Low	Low	High*
IgA	Low	Low	Low
IgE	Low	Low	Normal
On IvIg treatment	Yes	Yes	No
EBV antibodies	N/A (DNA negative)	N/A	High
RhF	N/A	N/A	N/A
dsDNA antibodies	Negative	N/A	N/A
ANCA	Positive (pANCA)‡	N/A	N/A
ANA	Negative	N/A	Negative

Supplementary Materials

Supplementary Text

Materials and Methods

Supplementary Figures 1-10

Supplementary Tables 1-5

Supplementary Movies 1-4

Supplementary References (*1-31*)

References

1. Bousfiha, A. *et al.* The 2015 IUIS Phenotypic Classification for Primary Immunodeficiencies. *J Clin Immunol* **35,** 727–738 (2015).
2. Picard, C. *et al.* Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *J Clin Immunol* **35,** 696–726 (2015).
3. Arason, G. J., Jorgensen, G. H. & Ludviksson, B. R. Primary immunodeficiency and autoimmunity: lessons from human diseases. *Scand J Immunol* **71,** 317–328 (2010).
4. Notarangelo, L. D. Primary immunodeficiencies. *J Allergy Clin Immunol* **125,** S182–94 (2010).
5. Conley, M. E. & Casanova, J.-L. Discovery of single-gene inborn errors of immunity by next generation sequencing. *Curr Opin Immunol* **30,** 17–23 (2014).
6. Deau, M.-C. *et al.* A human immunodeficiency caused by mutations in the PIK3R1 gene. *J Clin Invest* **125,** 1764–1765 (2015).
7. Lo, B. *et al.* Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy. *Science* **349,** 436–440 (2015).
8. Cunningham-Rundles, C. The many faces of common variable immunodeficiency. *Hematology Am Soc Hematol Educ Program* **2012,** 301–305 (2012).
9. Rieux-Laucat, F. & Casanova, J.-L. Immunology. Autoimmunity by haploinsufficiency. *Science* **345,** 1560–1561 (2014).
10. Lo, B. *et al.* CHAI and LATAIE: new genetic diseases of CTLA-4 checkpoint insufficiency. *Blood* **128,** 1037–1042 (2016).
11. Vahedi, G. *et al.* STATs shape the active enhancer landscape of T cell populations. *Cell* **151,** 981–993 (2012).
12. Whyte, W. A. *et al.* Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* **153,** 307–319 (2013).
13. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153,** 320–334 (2013).
14. Vahedi, G. *et al.* Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* **520,** 558–562 (2015).
15. Roychoudhuri, R. *et al.* BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. *Nature* **498,** 506–510 (2013).
16. Igarashi, K., Ochiai, K., Itoh-Nakadai, A. & Muto, A. Orchestration of plasma cell differentiation by Bach2 and its gene regulatory network. *Immunol Rev* **261,** 116–125 (2014).
17. Ferreira, M. A. R. *et al.* Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378,** 1006–1014 (2011).
18. Cooper, J. D. *et al.* Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat Genet* **40,** 1399–1401 (2008).
19. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42,** 1118–1125 (2010).
20. Dubois, P. C. A. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* **42,** 295–302 (2010).
21. Jin, Y. *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat Genet* **44,** 676–680 (2012).
22. International Multiple Sclerosis Genetics Consortium *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476,** 214–219 (2011).
23. Nakayama, Y. *et al.* A limited number of genes are involved in the differentiation of germinal center B cells. *J. Cell. Biochem.* **99,** 1308–1325 (2006).
24. Ochiai, K. *et al.* Plasmacytic transcription factor Blimp-1 is repressed by Bach2 in B cells. *J Biol Chem* **281,** 38226–38234 (2006).
25. Muto, A. *et al.* The transcriptional programme of antibody class switching involves the repressor Bach2. *Nature* **429,** 566–571 (2004).
26. Kuwahara, M. *et al.* The Menin-Bach2 axis is critical for regulating CD4 T-cell senescence and cytokine homeostasis. *Nat Commun* **5,** 3555 (2014).
27. Povoleri, G. A. M. *et al.* Thymic versus induced regulatory T cells - who regulates the regulators?

Front. Immunol. **4,** 169 (2013).
28. Rosbrook, G. O., Stead, M. A., Carr, S. B. & Wright, S. C. The structure of the Bach2 POZ-domain dimer reveals an intersubunit disulfide bond. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 26–34 (2012).
29. Uhlig, H. H. *et al.* The diagnostic approach to monogenic very early onset inflammatory bowel disease. *Gastroenterology* **147,** 990–1007.e3 (2014).
30. Deane, S., Selmi, C., Naguwa, S. M., Teuber, S. S. & Gershwin, M. E. Common variable immunodeficiency: etiological and treatment issues. *Int. Arch. Allergy Immunol.* **150,** 311–324 (2009).
31. Salzer, U. & Grimbacher, B. Monogenetic defects in common variable immunodeficiency: what can we learn about terminal B cell differentiation? *Curr Opin Rheumatol* **18,** 377–382 (2006).
32. Iwata, M. *et al.* Retinoic acid imprints gut-homing specificity on T cells. *Immunity* **21,** 527–538 (2004).
33. Cassani, B. *et al.* Gut-Tropic T Cells That Express Integrin α4β7 and CCR9 Are Required for Induction of Oral Immune Tolerance in Mice. *Gastroenterology* **141,** 2109–2118 (2011).
34. Igarashi, K., Ochiai, K. & Muto, A. Architecture and dynamics of the transcription factor network that regulates B-to-plasma cell differentiation. *J Biochem* **141,** 783–789 (2007).
35. Seidman, J. G. & Seidman, C. Transcription factor haploinsufficiency: when half a loaf is not enough. *J Clin Invest* **109,** 451–455 (2002).
36. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
37. Qian, J. *et al.* B Cell Super-Enhancers and Regulatory Clusters Recruit AID Tumorigenic Activity. *Cell* **159,** 1524–1537 (2014).
38. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* **6,** e1001154 (2010).
39. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
40. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107,** 21931–21936 (2010).
41. Khan, A. & Zhang, X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res* **44,** D164–71 (2016).
42. Parker, S. C. J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci* **110,** 17921–17926 (2013).
43. Roychoudhuri, R. *et al.* BACH2 regulates CD8(+) T cell differentiation by controlling access of AP-1 factors to enhancers. *Nat Immunol* **17,** 851–860 (2016).
44. Shinnakasu, R. *et al.* Regulated selection of germinal-center cells into the memory B cell compartment. *Nat Immunol* **17,** 861–869 (2016).
45. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).

Figure 1

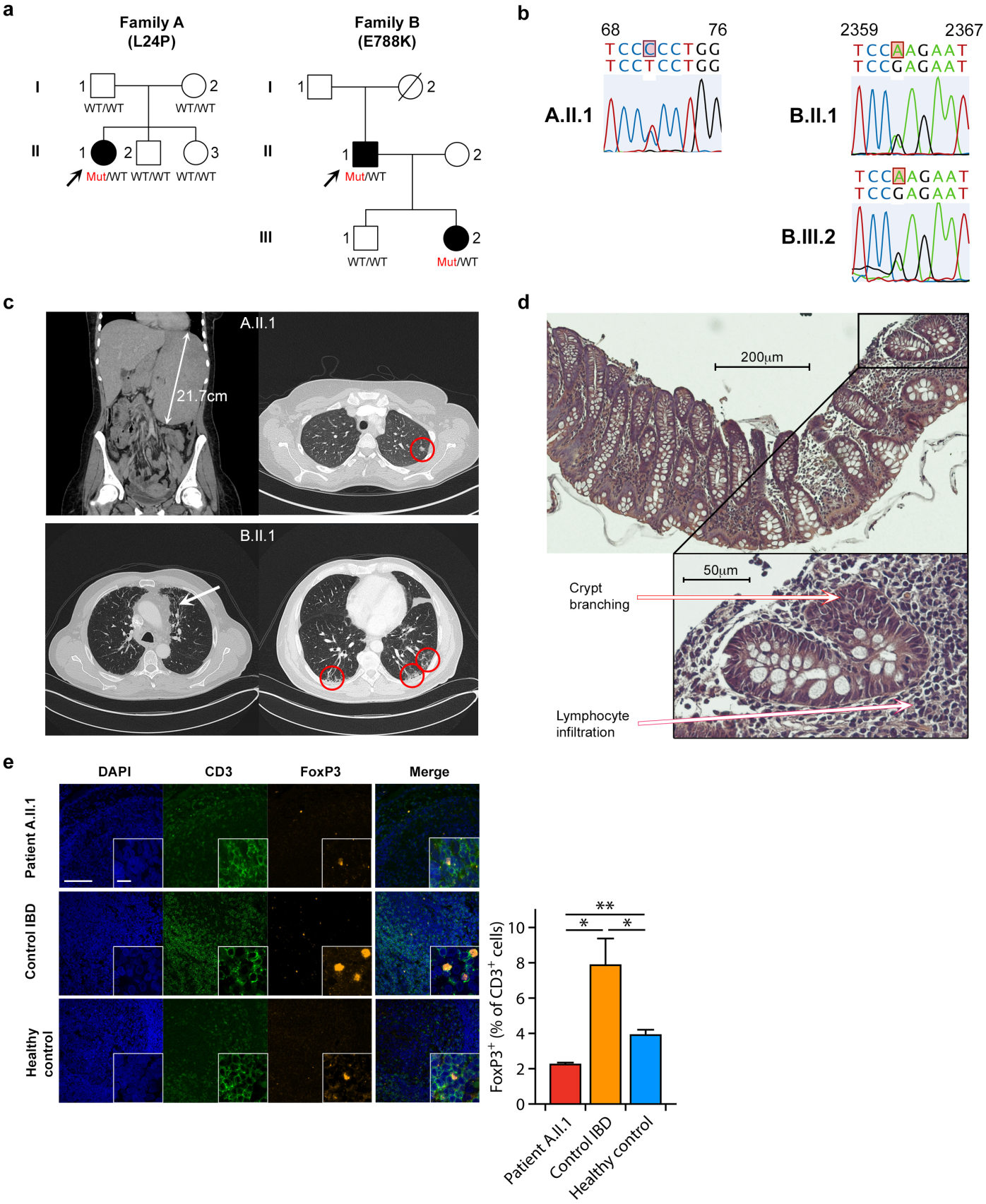

a
Family A
(L24P)
Family B
(E788K)
I
II
III
WT/WT
Mut/WT
b
68
76
TCCCCCTGG
TCCTCCTGG
A.II.1
2359
2367
TCCAAGAAT
TCCGAGAAT
B.II.1
B.III.2
c
A.II.1
21.7cm
B.II.1
d
200μm
50μm
Crypt
branching
Lymphocyte
infiltration
e
DAPI
CD3
FoxP3
Merge
Patient A.II.1
Control IBD
Healthy
control
FoxP3+ (% of CD3+ cells)
Patient A.II.1
Control IBD
Healthy control

Figure 2

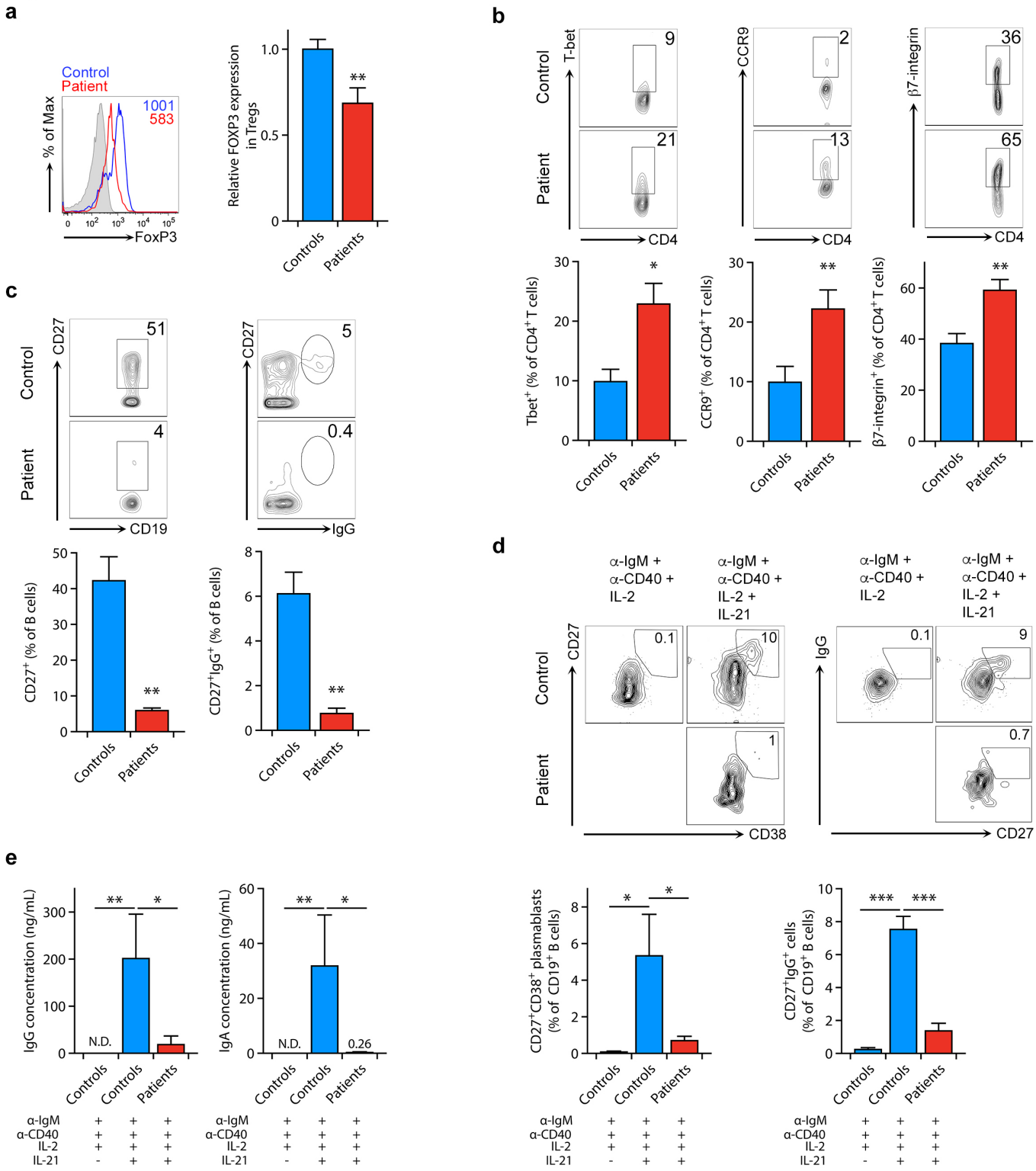

a
Control
Patient
1001
583
% of Max
FoxP3
Relative FOXP3 expression in Tregs
Controls
Patients
b
Control
Patient
T-bet
CD4
CCR9
β7-integrin
Tbet+ (% of CD4+ T cells)
CCR9+ (% of CD4+ T cells)
β7-integrin+ (% of CD4+ T cells)
c
CD27
CD19
IgG
CD27+ (% of B cells)
CD27+IgG+ (% of B cells)
d
α-IgM + α-CD40 + IL-2
α-IgM + α-CD40 + IL-2 + IL-21
CD38
e
IgG concentration (ng/mL)
IgA concentration (ng/mL)
N.D.
0.26
CD27+CD38+ plasmablasts (% of CD19+ B cells)
CD27+IgG+ cells (% of CD19+ B cells)
α-IgM
α-CD40
IL-2
IL-21

Figure 3

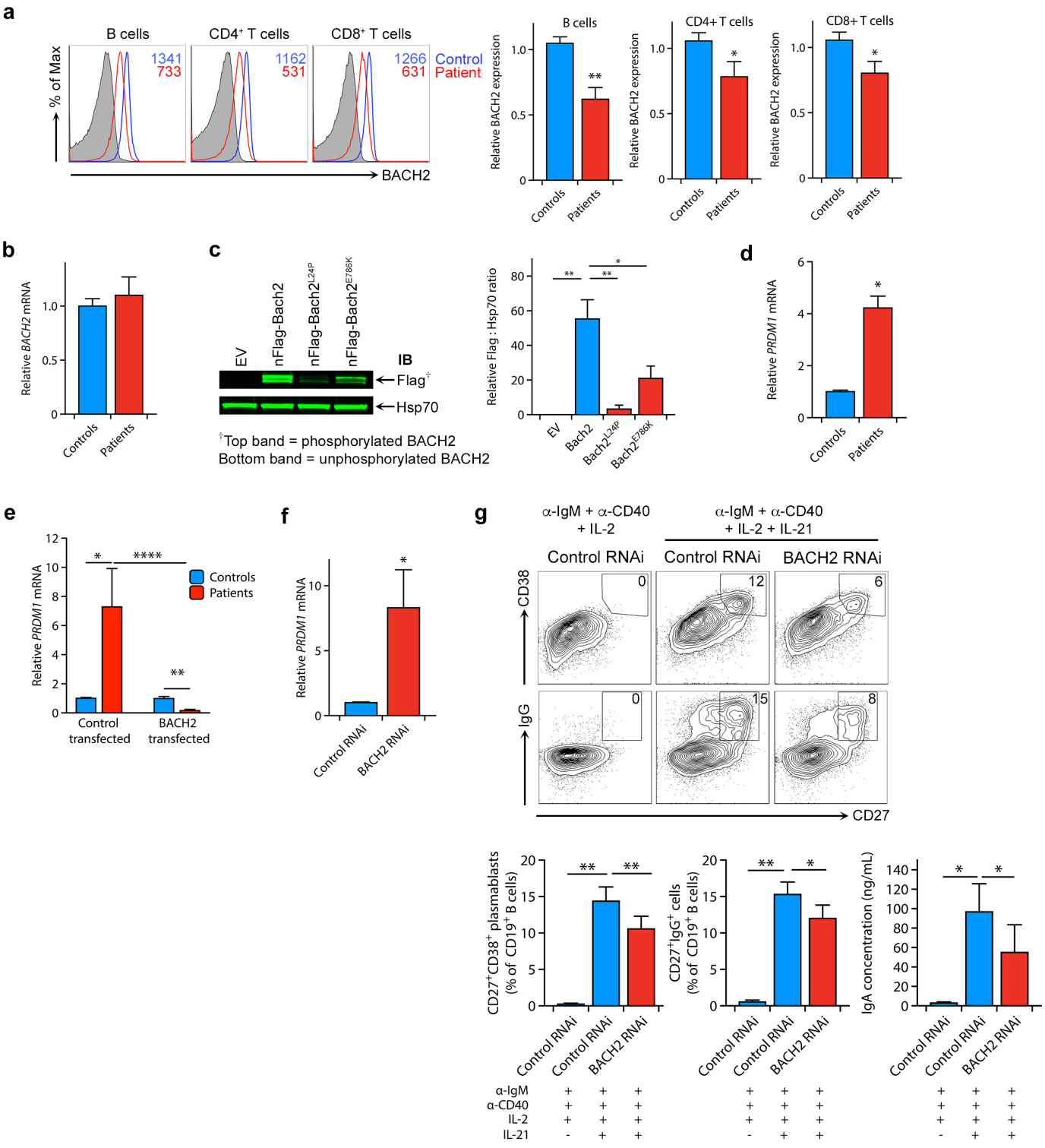

a
B cells
CD4⁺ T cells
CD8⁺ T cells
% of Max
1341
733
1162
531
1266
631
Control
Patient
BACH2
Relative BACH2 expression
Controls
Patients
CD4+ T cells
CD8+ T cells
b
Relative BACH2 mRNA
c
EV
nFlag-Bach2
nFlag-Bach2L24P
nFlag-Bach2E786K
IB
Flag+
Hsp70
†Top band = phosphorylated BACH2
Bottom band = unphosphorylated BACH2
Relative Flag : Hsp70 ratio
Bach2
Bach2L24P
Bach2E786K
d
Relative PRDM1 mRNA
e
Controls
Patients
Control transfected
BACH2 transfected
f
Control RNAi
BACH2 RNAi
g
α-IgM + α-CD40 + IL-2
α-IgM + α-CD40 + IL-2 + IL-21
Control RNAi
BACH2 RNAi
CD38
IgG
CD27
CD27+CD38+ plasmablasts (% of CD19+ B cells)
CD27+IgG+ cells (% of CD19+ B cells)
IgA concentration (ng/mL)
α-IgM
α-CD40
IL-2
IL-21

Figure 4

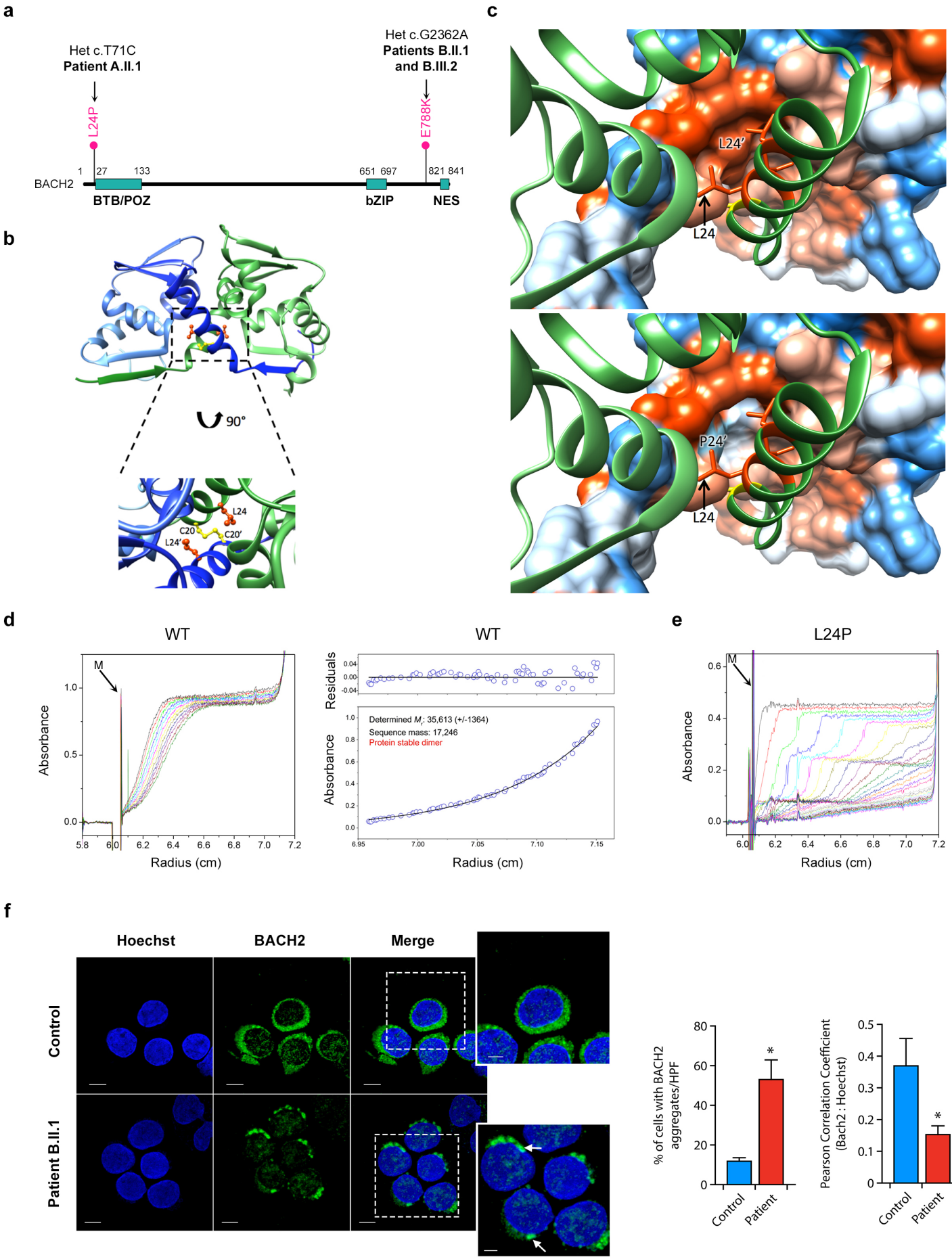

a
Het c.T71C
Patient A.II.1
Het c.G2362A
Patients B.II.1
and B.III.2
L24P
E788K
BACH2
1
27
133
651
697
821
841
BTB/POZ
bZIP
NES
b
90°
L24
C20
C20'
L24'
c
L24'
L24
P24'
L24
d
WT
WT
M
Absorbance
Residuals
Determined M: 35,613 (+/-1364)
Sequence mass: 17,246
Protein stable dimer
Radius (cm)
e
L24P
M
Absorbance
Radius (cm)
f
Hoechst
BACH2
Merge
Control
Patient B.II.1
% of cells with BACH2 aggregates/HPF
Pearson Correlation Coefficient (Bach2 : Hoechst)
Control
Patient

Figure 5

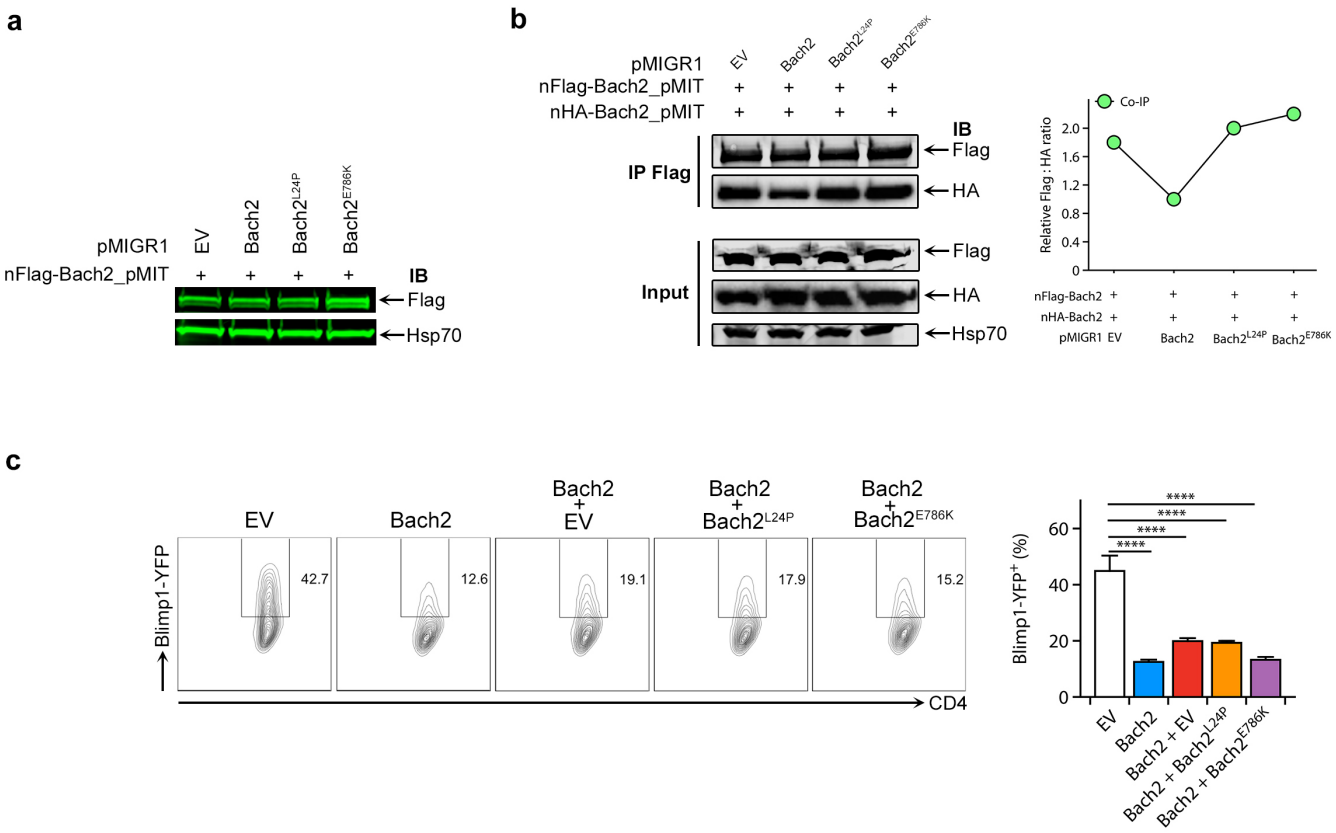

a
pMIGR1
EV
Bach2
Bach2L24P
Bach2E786K
nFlag-Bach2_pMIT
+
IB
Flag
Hsp70
b
pMIGR1
EV
Bach2
Bach2L24P
Bach2E786K
nFlag-Bach2_pMIT
nHA-Bach2_pMIT
IB
IP Flag
Flag
HA
Input
Flag
HA
Hsp70
Co-IP
Relative Flag : HA ratio
2.0
1.6
1.2
0.8
0.4
0
nFlag-Bach2
nHA-Bach2
pMIGR1 EV
Bach2
Bach2L24P
Bach2E786K
c
EV
Bach2
Bach2 + EV
Bach2 + Bach2L24P
Bach2 + Bach2E786K
Blimp1-YFP
42.7
12.6
19.1
17.9
15.2
CD4
Blimp1-YFP+ (%)
60
40
20
0
****
****
****
****
EV
Bach2
Bach2 + EV
Bach2 + Bach2L24P
Bach2 + Bach2E786K

Figure 6

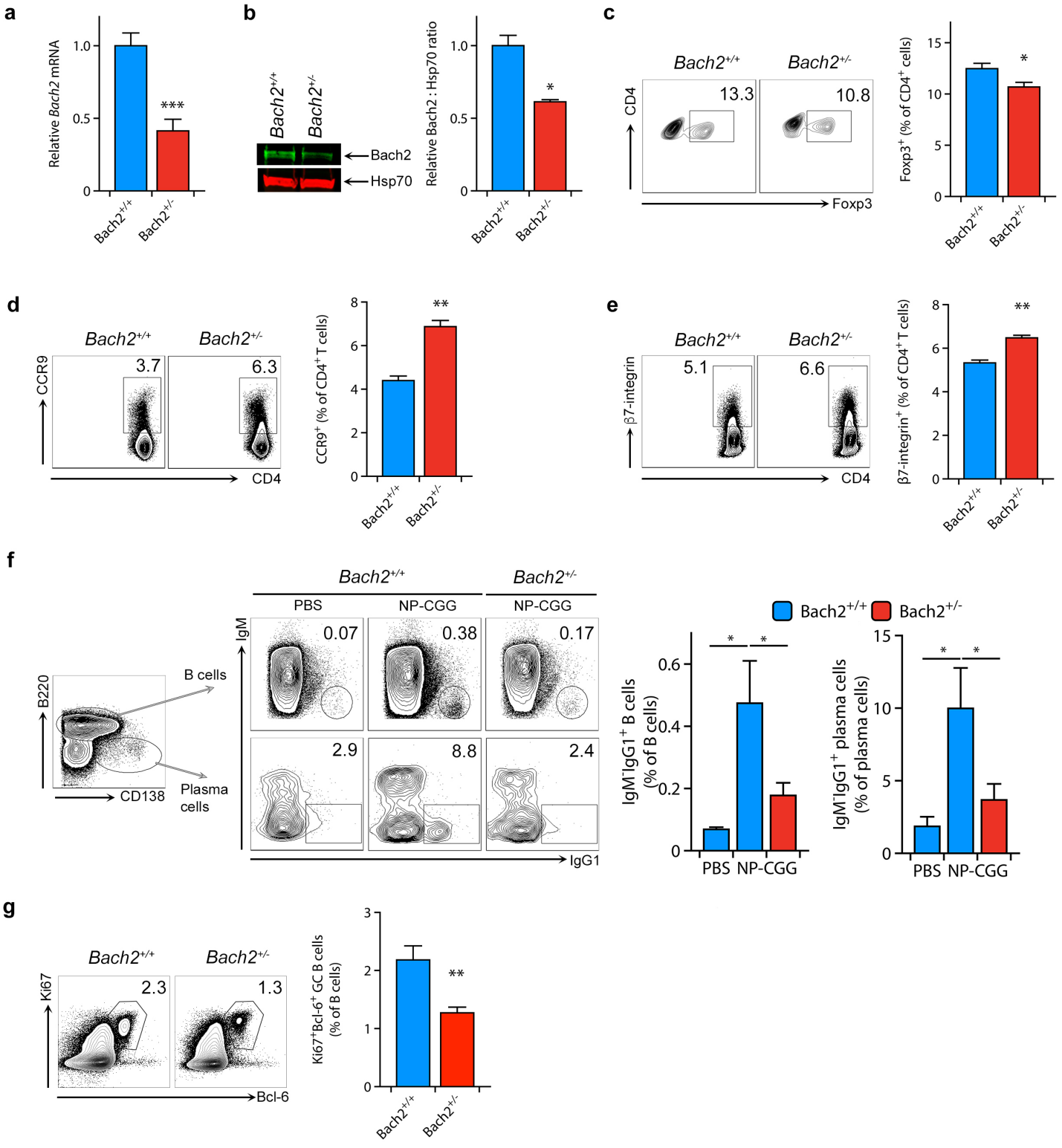
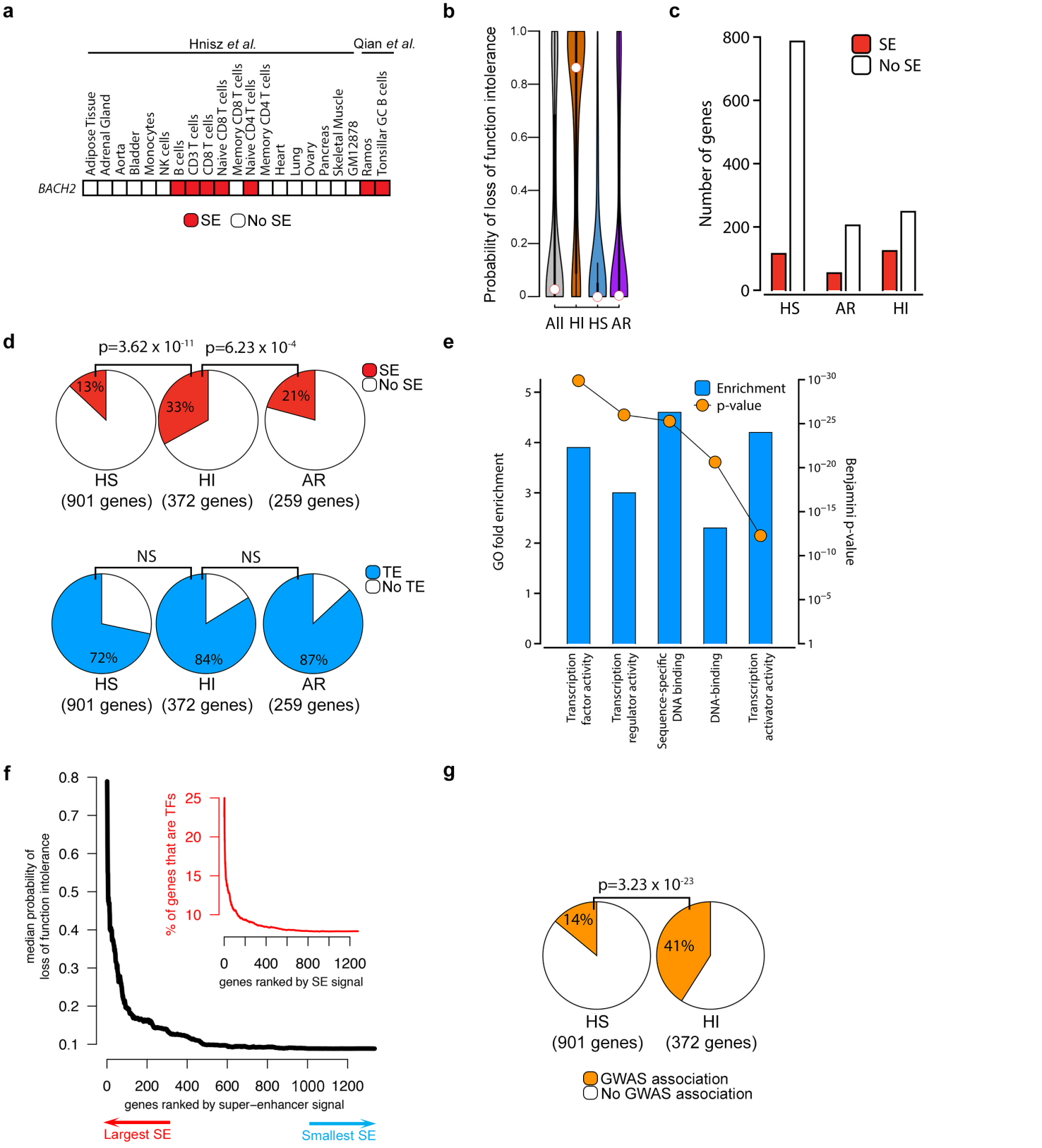

a
Relative Bach2 mRNA
1.0
0.5
0
***
Bach2+/+
Bach2+/-
b
Bach2+/+
Bach2+/-
Bach2
Hsp70
Relative Bach2 : Hsp70 ratio
1.0
0.5
0
*
Bach2+/+
Bach2+/-
c
Bach2+/+
Bach2+/-
CD4
13.3
10.8
Foxp3
Foxp3+ (% of CD4+ cells)
15
10
5
0
*
Bach2+/+
Bach2+/-
d
Bach2+/+
Bach2+/-
CCR9
3.7
6.3
CD4
CCR9+ (% of CD4+ T cells)
8
6
4
2
0
**
Bach2+/+
Bach2+/-
e
Bach2+/+
Bach2+/-
β7-integrin
5.1
6.6
CD4
β7-integrin+ (% of CD4+ T cells)
8
6
4
2
0
**
Bach2+/+
Bach2+/-
f
Bach2+/+
Bach2+/-
PBS
NP-CGG
NP-CGG
B220
B cells
CD138
Plasma cells
IgM
0.07
0.38
0.17
2.9
8.8
2.4
IgG1
Bach2+/+
Bach2+/-
IgM-IgG1+ B cells (% of B cells)
0.6
0.4
0.2
0
* *
PBS NP-CGG
IgM-IgG1+ plasma cells (% of plasma cells)
15
10
5
0
* *
PBS NP-CGG
g
Bach2+/+
Bach2+/-
Ki67
2.3
1.3
Bcl-6
Ki67+Bcl-6+ GC B cells (% of B cells)
3
2
1
0
**
Bach2+/+
Bach2+/-

Figure 7


a
Hnisz et al.
Qian et al.
Adipose Tissue
Adrenal Gland
Aorta
Bladder
Monocytes
NK cells
B cells
CD3 T cells
CD8 T cells
Naive CD8 T cells
Memory CD8 T cells
Naive CD4 T cells
Memory CD4 T cells
Heart
Lung
Ovary
Pancreas
Skeletal Muscle
GM12878
Ramos
Tonsillar GC B cells
BACH2
SE
No SE
b
Probability of loss of function intolerance
All HI HS AR
c
Number of genes
SE
No SE
HS
AR
HI
d
p=3.62 x 10-11
p=6.23 x 10-4
13%
33%
21%
SE
No SE
HS
(901 genes)
HI
(372 genes)
AR
(259 genes)
NS
NS
72%
84%
87%
TE
No TE
e
GO fold enrichment
Enrichment
p-value
Benjamini p-value
Transcription factor activity
Transcription regulator activity
Sequence-specific DNA binding
DNA-binding
Transcription activator activity
f
median probability of loss of function intolerance
% of genes that are TFs
genes ranked by SE signal
genes ranked by super−enhancer signal
Largest SE
Smallest SE
g
p=3.23 x 10-23
14%
41%
HS
(901 genes)
HI
(372 genes)
GWAS association
No GWAS association